## REMARKS:

This paper is herewith filed in response to the Examiner's Office Action mailed on October 25, 2007 for the above-captioned U.S. Patent Application. This office action is a rejection of claims 1-46 of the application.

More specifically, the Examiner has rejected claims 1, 4, 7, 9-15, and 17 under 35 USC 102(b) as being anticipated by Garfield (Doctoral Dissertation); rejected claims 1, 4, 7, 9-15, 17, 19, 22, 25, 27-33, 35, 37, and 40 under 35 USC 103(a) as being unpatentable over Friedman (US6,182,029) in view of Brecher (US7,054,754) and in view of Moore (US5,577,239); rejected claims 2-3, 6, 20-21, 24, 38-39, and 42 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher and in view of Moore and in further view of Hull (US6,332,138); rejected claims 18 and 36 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher and in view of Moore and further in view of Kemp (Chem. Inf. Comput. Sci., Vol. 38); rejected claims 5, 16, 23, 34, and 41 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher and in view of Moore and further in view of Dittmar (J. Chem. Inf. Comput. Sci. Vol. 23, No. 3); rejected claims 8 and 26 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher and in view of Moore and in further view Drefahl (J. Chem. Inf. Comput. Sci., Vol. 33); rejected claim 43 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher and in view of Moore and Shivarati (Computer December 1992); and rejected claim 44 under 35 USC 103(a) as being unpatentable over Friedman in view of Brecher in view of Moore and Shivarati and further in view of Leiter (J. Chem. Doc., Vol. 15, 1965).

The Applicants respectfully traverse the rejections.

Claims 1-3, 6, 8, 19-21, 24, 26, 37-39, 42-44, and 46 have been amended. Claims 4-5, 22-23, and 40-41 have been canceled. Support for the amendments can be found at least on page 9 lines 5-11, page 11 lines 18-20, page 14 lines 2-18, page 15 lines 15-17, and page 16, lines 1-4. No new matter is added.

Regarding the rejection of claim 1 under 35 USC 102(b) the Applicants note that claim 1 has been amended to recite:

> A method to process a text document, comprising: partitioning text of the text document and assigning semantic meaning to words of the partitioned text, where assigning comprises applying a plurality of regular expressions, rules and a plurality of dictionaries to recognize chemical name fragments; recognizing any substructures present in the chemical name fragments; determining structural connectivity information of the chemical name fragments and recognized substructures; extracting information associated with the recognized chemical name fragments and substructures of the text document and indexing the extracted information in a text index; indexing representations of the chemical name fragments and the recognized substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables; storing the text index in association with the indexed representations in a searchable index; and providing a graphical user interface to search the index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in the representation form, where the entering is by at least one of text form or graphical selection.

The Applicants note that claim 1 has been amended to incorporate features similar to claim 4 and claim 5.

In the Office Action the Examiner states:

> "Regarding claims 4, Garfield teaches searching the index by fragment or substructure name and/or connectivity (p. 466, Relationship Between Nomenclature and Searching)."

Here Garfield discloses:

> "A by-product of this study is the clearer understanding of the relationship between nomenclature and chemical searching requirements When the computer analysis of the chemical name is completed, the parsed expression that results from the analysis could be used by the computer to perform very adequate generic as well as specific searches. **If the chemists specifies the type of chemical in which he is interested in terms of morphemes instead of conventional chemical class names, generic searches become quite simple**. Hence, a search for all hexenols becomes a search for all chemicals which contain the morpheme co-occurrence hexen and the morpheme ol. If he is interested in any six-carbon-chain-alcohol he need only specify the presence of hex and ol, where hex must be

> the carbon containing morpheme, not the multiplier morpheme as in hexa-chlorooctane," (emphasis added), (page 466, lines 27-36).

The Applicants submit that Garfield clearly does not disclose or suggest at least where claim 1 recites in part "**indexing the extracted identifying information in a text index; indexing representations of the chemical name fragments and the recognized substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables; storing the text index in association with the indexed representations in a searchable index**." As cited Garfield is merely seen to disclose that a search of a type of chemical can be made more specific where the search is performed using morphemes instead of conventional chemical class names. Moreover, Garfield clearly can not be seen to relate to providing **a graphical user interface** to search the index as in claim 1.

For at least the reasons stated the Applicants contend that the rejection of claim 1 under 35 USC 102(b) over Garfield is overcome and the rejection should be removed.

Regarding the rejection of claim 1 under 35 USC 103(a) the Applicants traverse the rejection.

In the Office Action the Examiner states:

> "Friedman shows a method and system for extracting information from natural language text data. Friedman shows information is extracted from text documents (col. 4, line 59-63). Friedman shows that the text of the text document is partitioned into phrases (col. 6, line 36-45). Friedman shows that partitioned phrases are further parsed to assign semantic meaning to words (col. 6, line 63-65). **Friedman suggests that chemical information can be identified and extracted (col. 11, line 34-50)**. Friedman shows the method provides reliable and efficient access to information within a document and is useful for retrieving and summarizing relevant information in documents (col. 4, line 59-67)," (emphasis added).

Firstly, the Applicant submit the Examiner's is clearly in error where the Examiner appears to assert that Friedman suggests the chemical information can be identified and extracted from other

information in a text document such as in the invention.

To support this apparent assertion the Examiner cites where Friedman discloses:

> "FIG. 3 shows a block diagram of a second embodiment of the information extraction (MedLEE) program of FIG. 1. The modified program 300 includes **a tagger routine 16 for linking the structured output described previously with respect to FIG. 1** to the corresponding words in the original sentences of the text data input. Preferably, the **tagger 16 utilizes markup languages**, such as Hypertext Markup Language (HTML) and Extensible Markup Language (XML), which are derived from Standard Generalized Markup Language (SGML) and which are used rendering documents for the World Wide Web. Widespread adoption of markup languages are evidenced by: the Text Encoding Initiative (TEI) which uses SGML to encode literature; **Chemical Markup Language (CML), which involves documentation of chemical compounds using SGML**; and Open Financial Exchange (OFE), which is an SGML standard format for interchange of financial transactions," (col. 11, lines 34-50)," (emphasis added).

The Applicants submit that as cited by the Examiner Friedman appears to disclose that the tagger routine 16 is using Chemical Markup Language (CML) to merely **link**, for the World Wide Web, chemical compound words which appear to be part of a **previous output** in Friedman. The Applicants can not find in all of Friedman where chemical compound names are identified in a text document and partitioned separately from other words. Clearly, a routine for **linking** with a markup language can not be seen to disclose or suggest identifying and extracting chemical compounds out of a text document as appears to be indicated by the Examiner to support the rejection of claim 1.

Further, Friedman discloses:

> "The preprocessor 11 thus performs lexical lookup to **identify and categorize multi-word and single word phases within each sentence**. The output of this component consists of a list of word positions where each position is associated with a word or multi-word phrase in the report. For example, assuming that the sentence "spleen appears to be moderately enlarged" is at the beginning of the report, it would be represented as the list where position 1 is associated with "spleen", position 2 with the multi-word phrase "appears to be", position 5 with "moderately", and 6 with "enlarged". The remainder of the list of word positions

> would be associated with the remaining words in the report," (emphasis added), (col. 7, lines 17-29).

The Applicants argue that Friedman appears to identify and categorize all words and phrases in a sentence. The Applicants can not find in Friedman where a text document is partitioned and the partitions examined to identify and extract chemical terms from other terms in the text document as at least appears to be required to support the rejection of claim 1.

Further in the rejection of claim 1 the Examiner states:

> **Friedman does not show the application of regular expressions and a plurality of chemical dictionaries to recognize chemical *names*** or storing information in a searchable index.
>
> "Brecher shows a method system and computer program product for **processing text documents to extract chemical information**. Brecher shows the application of regular expression (col. 5, line 41-45) and a plurality of dictionaries to recognize chemical names (col. 6, line 29-40). Brecher shows that the lexicon has at least a sub lexicon to identify stopwords (col. 8, line 49-50), prefixes (001. 9, line 55) or suffixes (col. 11, line 43). Brecher shows that substructures are recognized (col. 6, line 31-33). Brecher shows that structural connectivity is determined (col.7, line 35-57). Brecher et al. shows that identifying information is extracted from the substructures and fragments to produce a fully parsed chemical name that is correlated to a chemical structure. Brecher shows the method allows chemical names to be accurately converted to chemical structures in real time or in nearly real time to provide users with a powerful, practical tool (col. 2, line 11-14)," (emphasis added).

Firstly, the Applicants disagree with the Examiner's apparent assertion the Brecher can be seen to extract chemical information from an output in Friedman. The Applicants contend that this apparent assumption is clearly not the case. Further, the Applicants submit that the Examiner is unclear in the rejection as to whether it is Friedman or Brecher who is allegedly performing a function of identifying and extracting chemical information apart from other information in a text document. The Applicants respectfully request that clarification is provided in a non-final office action. The Applicants contend that neither Friedman nor Brecher can be seen to disclose or suggest identifying and extracting only chemical information apart from all other non-chemical words.

Brecher discloses:

> "**A chemical name 12 is supplied** via one or more input systems such as end-user keyboard input 14, file-based input 16, or World-Wide Web query input 18," (emphasis added), (col. 2, lines 47-50).

As can be seen in Brecher a **chemical name is supplied**. The Applicants submit that Brecher can not be combined with Friedman for at least the reason that Friedman does not relate to partitioning a text document to identify chemical names from other words or phrases in the text document. The Applicants contend that the rejection is improper at least for the reason that an inventive step is clearly missing. The Applicant contends that Friedman would at least have to identify and extract the "multi-word and single word phases within each sentence" which comprise chemical names for other words in a text document and then provide the extracted chemical names to Brecher.

Moreover, for at least the reasons already stated neither Friedman nor Brecher can be seen to disclose or suggest "extracting **information associated with the recognized chemical name fragments and substructures** of the text document and indexing the extracted information in a text index," as in claim 1.

Clearly, a person skilled in the art would not be motivated to combine Friedman and Brecher. The Applicants contend that for at least these reasons such a combination would not be proper.

The Applicants contend that for at least the reasons already stated the rejection of claim 1 is seen to fail and the rejection should be removed.

Further, regarding the rejection of claim 1 the Examiner states:

> "Moore et al. shows a method of storing extracted identifying information in a searchable index (col. 4, line 28-35). Moore et al. shows that the index can be searched by a combination of substructure names (col. 7, line 47-48) and connectivities (col. 10, line 43-46). Moore et al. shows the method has the

> advantage of simplified search queries (col. 12, line 42-46). Moore et al. shows the method has the further advantages of reducing database development and maintenance costs, simplify interfacing with other information systems (col. 2, line 10-23)."

The Applicants note that Moore make no reference to providing a graphical user interface to search the index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in the representation form, where the entering is by at least one of text form or graphical selection as in claim 1. Further, Moore appears to disclose that the substructure search is performed in the same manner in which a search for duplicates is performed. The Applicants submit that Moore appears to disclose simply searching for a match of the search term.

In addition, Moore discloses:

> "Three cases are possible.
> Case 1. The query is a substructure of the retrieval structure. Note that Q1 (FIG. 5) is a substructure of S1 (FIG. 1). In this case, note that each Q1 key matches an S1 key (Q1 keys 1, 2, 3, 4 match S1 keys 1, 2, 3, 4 respectively).
>
> Case 2. The chosen key may retrieve a structure for which the query is not a substructure. For example, Q1 key 4 would match S2 key 4 even though Q1 is not a substructure of S2. However, the ABAM would eliminate this structure.
>
> Case 3. The chosen query key does not match any of the keys of a particular structure. For example, Q1 key 1 does not match any of the keys of S2. Therefore, S2 is eliminated as a match and ABAM is avoided," (col. 9, lines 45-57).

Thus, Moore discloses that searches of keys can be modified to search based either on an inclusion or exclusion of a search term in a key. However, Moore does not disclose or suggest that a search is performed using a combination of one or more chemical fragment names and one or more substructures as in claim 1.

Further, the Applicants note that in Moore there are also disclosed dynamic queries and structure

class searches. The dynamic queries appear to allow a user to store a search result as dynamic such that a user will be notified if a new substructure, chemical name, or molecular formula is added which matches the earlier stored dynamic search, (col. 11, lines 46-63). Further, as a method to reduce overhead Moore discloses that a structure class search can be used to limit the scope of searches to a class of chemical types (col. 12, lines 37-39). The Applicants contend that neither of these methods in Moore relate to a search performed by entering one or more chemical fragment names and entering one or more substructures in the representation form, where the entering is by at least one of text form or graphical selection as in claim 1.

The Applicants contend that for at least the reasons stated Moore can not be seen to disclose or suggest at least where claim 1 recites "providing a graphical user interface to search the index, where searching comprises entering one or more chemical fragment names and selecting one or more substructures."

Thus, for at least these reasons already stated the Applicants respectfully request that the Examiner remove the rejection of claim 1 and allow claim 1 to issue.

Furthermore, the Applicants submit that none of the references cited by the Examiner are seen to overcome a shortfall of Friedman, Brecher, Moore and Garfield as stated above.

Hull relates to a method to a method for computing chemical similarities using textual and chemical descriptors, (abstract). Hull discloses that two data bases are generated for a comparison and that these data bases are called $TIMI_T$ and $TIMI_C$. Further, Hull appears to indicate that these two data bases are combined in another database called $TIMI_{TC}$. In addition, Hull discloses that the "T" stands for text and "C" stands for chemistry, (col. 12, lines 44-53).

Hull discloses:

> "Finally, we can perform one special search in $TIMI_{TC}$ that can not be performed in either $TIMI_T$ or $TIMI_C$ individually--a combined structure and text query.

> **Combining both query types is advantageous because one can "tweak" a structural search with carefully chosen keywords**. For example, suppose the user is most interested in the possibility of toxicity with a given compound. **She can then add terms related to toxicity to the structural query, thereby ranking documents which discuss toxicity issues more highly**," (emphasis added), (col. 16. lines 23-32).

As disclosed in Hull, keywords are applied to "tweak" a structural search. The Applicants note that the example in Hull merely appears to indicate that descriptor terms i.e. toxicity are used to narrow the search and the resulting documents retrieved are ranked based on the terms. The Applicants contend that Hull can not be seen to relate to partitioning text of the text document in order to recognize substructures and chemical name fragments, and performing a search using a combination of **chemical fragment names and substructures** as in claim 1.

Dittmar relates to using a CAS online system to search chemical structures and nomenclature data, (page 93, col. 1, introduction). As cited by the Examiner Dittmar discloses "In the CAS ONLINE file-building procedure, the creation of substructure search screens for a substance begins with a fragment generation step, in which structure fragments and other search data are generated from the Registry connection table of substance," (page 99, col. 2, para. 2, sentence 1).

Further Dittmar discloses "The CAS ONLINE system allows a search query to be defined in terms of one or several structure diagrams and / or manually encoded screen sets, with components combined by AND, OR, and NOT Boolean operators," (page 95, col. 1, 2nd paragraph, lines 1-4). The Applicants can find nothing in Dittmar to can not be seen to relate to **partitioning text of the text document in order to recognize substructures and chemical name fragments**, and performing a search using a combination of **chemical fragment names and substructures** as in claim 1.

Further, the Applicants note that independent claims 19, 37, and 43 have been amended to include features similar to claim 1. Thus, for at least the reasons already stated the rejection of all claims 1, 19, 37, and 43 are seen to be overcome and the rejections of these claims should be removed.

For at least the reasons already stated the Applicants contend that the references cited are not seen to overcome the shortfalls as stated above. Further, even if the references were combined, which is not agreed is feasible or possible, for at least the reasons stated the combination would still not disclose or suggest the claims. Thus, the rejection of these claims should be removed.

Regarding the rejection of claim 8 the Examiner states:

> "Claims 8 and 26 are directed to a structure dictionary comprising at least one of a MOL dictionary and a SMILES dictionary. Friedman, in view of Brecher and in view of Moore et al. does not teach a structure dictionary comprising at least one of a MOL dictionary and a SMILES dictionary. Drefahl et al. teach a structure dictionary comprising at least one of a MOL dictionary and a SMILES dictionary (abstract, sent. 3)."

The Applicants note that claim 8 has amended to recite:

"A method as in claim 1, where the representations comprise MOL type representations and SMILES type representations."

The Applicants note that as cited by the Examiner Drefahl simply discloses "The database contains compounds that are represented by SMILES notations," (abstract, set. 3). Although the Applicants do not agree that Drefahl discloses or suggests a SMILES dictionary as in claim 8, the Applicants contend that Drefahl clearly does not disclose or suggest at least where claim 8 recites where **the representations comprise MOL type representations and SMILES type representations**.

For at least this reason the Applicants submit the rejection of claim 8 should be removed.

Further, as claims 26 and 46 have been amended to recite similar features of claim 8 the rejections should be removed for all claims 8, 26, and 46.

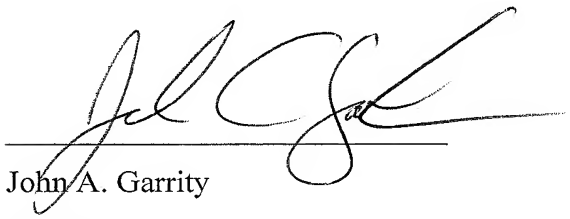
In addition, the Applicants respectfully submit that the Applicants do not acquiesce to any rejection made in the Office Action but not addressed in this Response.

In addition, for at least the reason that claims 2-3, and 6-18; claims 20-21, and 24-36; and claims 38-40, and 42; and claims 44-46 depend from independent claims 1, 19, 37, and 43, respectively, the references cited are not seen to disclose or suggest all these claims and the rejection of all these claims should be removed.

Based on the above explanations and arguments, it is clear that the references cited cannot be seen to disclose or suggest claims 1-3, 6-21, 24-40, and 42-46. The Examiner is respectfully requested to reconsider and remove the rejections of all claims and to allow all of the pending claims 1-3, 6-21, 24-40, and 42-46.

For all of the foregoing reasons, it is respectfully submitted that all of the claims now present in the application are clearly novel and patentable over the prior art of record. Should any unresolved issue remain, the Examiner is invited to call Applicants' attorney at the telephone number indicated below.

Respectfully submitted:



4/25/2008

John A. Garrity
Reg. No.: 60,470
Customer No.: 48237
HARRINGTON & SMITH, PC
4 Research Drive
Shelton, CT 06484-6212
Telephone: (203)925-9400
Facsimile: (203)944-0245

Date